Ontologies in the biomedical domain.

Part II: Examples

Anita Burgun[*], Olivier Bodenreider[**]

[*] *LIM- University of Rennes, Av. Pr Léon Bernard, 35043 Rennes, France*
[**] *U.S. National Library of Medicine, Bethesda, MD, USA*

Anita.Burgun@univ-rennes1.fr, olivier@nlm.nih.gov

Accepted for publication in JAMIA

Abstract

This paper is the second of two review articles devoted to ontologies in the biomedical domain. The focus of the first article is on the principles applied to building ontologies and we now examine examples of biomedical ontologies. A brief presentation of some currently available medical ontologies (OpenGALEN, UMLS Semantic Network, SNOMED CT, Digital Anatomist, and MENELAS ontology) is provided as well as the description of the biomedical domain in two general ontologies (OpenCyc and WordNet). Using the representation of Blood in each system as an example, we examine seven approaches to representing this concept. This study revealed issues in compatibility among the representations, partially related to the gap between expert knowledge and general knowledge. It also brought out the lack of shared definition for some core concepts of the biomedical domain.

I. Introduction

Ontology, as a branch of philosophy, is related to the study of what is. In a previous paper [1], we have presented some principles generally used in ontology design. In practice, however, several conceptualizations of a domain, often called “ontologies”, may be produced by different groups of researchers coming from, for example, database modeling or knowledge engineering. In this paper, we investigate how existing ontologies may provide different views of the biomedical domain. First, we examine the representation of biomedicine in general systems such as OpenCyc and WordNet. While OpenCyc aims at describing common sense knowledge about the world, WordNet is based on a linguistic approach. We then study three systems of concepts whose intended coverage is the whole biomedical domain, namely GALEN, the UMLS, and SNOMED CT. An ontology of anatomy, the Digital Anatomist Foundational Model, is also explored. Our rationale for focusing on this subdomain is that anatomy is central to biomedicine and essentially stable. Therefore, an ontology of anatomy should be sharable and present in virtually every biomedical ontology. Finally, as an example of application ontology, we examine the ontology developed as part of the MENELAS project.

After a brief presentation of the characteristics of these ontologies, we use the representation of the concept *Blood* in each system to illustrate their common features and differences. Issues in building a common, sharable framework for representing biomedical knowledge will be presented and discussed.

This study is a contribution to the Medical Ontology Research project currently being developed at the U.S. National Library of Medicine [2]. The major objective of this project is to develop methods whereby biomedical ontologies could be acquired from existing resources as well as validated against other knowledge sources. References for the systems presented in this paper are listed in Appendix A.

II. Representation of the biomedical domain in general ontologies

A. Representation of the biomedical domain in OpenCyc

Introduction

Cyc® is a general ontology that has been developed by Cycorp, Inc since 1984. This system is built upon a core of over 1,000,000 hand-entered assertions designed to capture a large portion of what people normally consider consensus knowledge about the world. The goal is to construct a foundation of basic “common sense” knowledge that will enable a variety of knowledge-intensive products and services. Cyc is intended to provide a deep layer of understanding that can be used by other programs to make them more flexible. The representation of knowledge in Cyc uses the formal language called CycL. The Cyc ontology consists of concepts and assertions which relate those concepts. Cyc allows the representation of "microtheories", each of which is essentially a bundle of assertions that share a common set of assumptions. Some microtheories focus on a particular domain of knowledge, a particular level of detail, a particular interval in time, etc. The microtheory mechanism allows Cyc to independently maintain assertions which are prima facie contradictory and enhances the performance of the Cyc system by focusing the inferencing process. OpenCyc™, the upper level, publicly available part of Cyc – currently in beta version – is expected to eventually comprise 6,000 concepts and 60,000 assertions about these concepts.

Ontological features

Specific to OpenCyc is the opposition between individuals (they may have parts or a structure, but do not have elements or subsets) and collections. In order to implement this feature, OpenCyc uses the following two structuring relations:

- *#$isa* means "is an instance of". `(#$isa El Col)` means that *El* is an element of the collection *Col*.
- *#$genls* is the relation between a collection and its superordinate. `(#$genls Col Sup)` means that *Sup* is a category that is a superordinate of *Col*.

As illustrated in Figure 1, *Thing*, the universal set, is the collection of everything. Using the concept *Cancer* as an example, we may examine the approach used for representing concepts in OpenCyc. *Cancer* is an instance of the type *Disease Type* `(#$isa #$Cancer #$DiseaseType)`. *Cancer* is a also a subordinate of *Ailment condition* `(#$genls #$Cancer #$AilmentCondition)`. Instances of *Ailment condition* are dynamic states of sickness, injuries, or physiological impairments. Comments are provided for concepts, e.g., "Cancer is characterized by abnormal (and usually rapid) growth of cells in some organ or system of the body; these growths are then prone to dispersal (metastasis) into other body regions". *Cancer* is also a subordinate of *Terminal physiological condition*, which means that "if left untreated, those ailments will certainly cause a patient to die – and may eventually do so despite treatment". Subsets of this collection include *Diabetes*, *Cancer*, *Botulism*, and *Rabies*.

In addition, *CancerFn* is a Cyc function that may be used to describe cancers according to the part or region of the animal's body in which they are found. Its argument is an animal body part and its result is a cancer. For example, `(#$CancerFn #$Throat)` represents the set of throat cancers, and is a subset of *Cancer*.

Additional knowledge may be represented within OpenCyc hierarchies. This knowledge corresponds to:

- generic concepts, e.g., path. Having *Path* as its superordinate, *Blood vessel* inherits the general properties of paths, such as origin and destination.
- properties, or attributes, e.g., symmetric (*Eye* is an instance of the collection *Symmetric anatomical part Type*), aquatic (*Fish* is an *Aquatic organism*).
- general common-sense knowledge, e.g., terminal condition. Making *Terminal physiological condition* a superordinate of *Cancer* conveys, through hierarchies, that cancer may lead to death.

General microtheories in OpenCyc include several microtheories of the biomedical domain, such as Biology knowledge or Ailment knowledge. Potential benefits of microtheories are twofold: (1) there exists assertions whose arguments are microtheories, e.g., "Everything true in Vertebrate Physiology knowledge is also true in Ailment knowledge", and (2) things can have distinct representations under distinct microtheories, e.g., in Animal Physiology, subordinates of *Sensor* include *Nose*, *Skin*, and *Ear*, while they include *Tactile sensor* and *Electromagnetic radiation sensor* in Naïve Physics.

B. Representation of the biomedical domain in WordNet

Introduction

WordNet® is an electronic lexical database that has been developed and maintained at Princeton University since 1985 [3]. WordNet organizes lexical information in terms of meanings and semantic relations. English nouns, verbs, adjectives and adverbs are organized into synonym sets, called synsets, each representing one underlying concept. Synset formation is based on the notions of synonymy (one meaning expressed by several word forms) and polysemy (word forms having several distinct meanings). Separate structures are built for each kind of linguistic items. For example the adjective "renal" and the noun "kidney", although representing the same meaning, belong to two distinct structures, and a specific relationship (pertainymy) relates the two forms. The current version (1.7) contains over 100,000 noun synsets. WordNet has been employed as a resource for many applications in natural language processing and information retrieval[*].

[*] http://www.cogsci.princeton.edu/~wn/papers/

Ontological features (noun hierarchy)

Several types of relations between synsets in the noun hierarchy are recorded in WordNet, including hyponymy (*is a*) and meronymy (*part of*). Each synset belongs to (at least) one *is-a* tree, and may additionally belong to several *part-of*-like trees. Hyponymy relations are instantiated between synsets, according to the following definition: A concept represented by the synset {x,x',…} is said to be a hyponym of the concept represented by the synset {y,y',…} if native speakers of English accept sentences constructed from such frames as "An x is a kind of y" [3].

The organization of the top level of the noun hierarchy essentially relies on empirical features: "In principle, it is possible to put some empty set at the top: to make {T} the hypernym of every synset that does not have a hypernym, thus pulling all nouns together into a single hierarchical structure. […] This device is sometimes convenient when using the hierarchy to estimate semantic distances, since a path can then be traced between any two words or any two synsets. The lexical justification is tenuous, however, because these abstract generic concepts carry so little semantic information; it is doubtful that people could agree on appropriate words to express them." [3:28].

Besides linguistics, WordNet has also been influenced by cognitive psychology. WordNet hierarchies reflect knowledge as perceived by most English speakers rather than systematically organized according to theoretical principles. WordNet divides the nouns into 11 hierarchies, each starting with a different "unique beginner". The list of unique beginners is given in Figure 2.

Several groups studied how knowledge is represented in WordNet. An ontological analysis of WordNet's top level is presented by Gangemi & al., and a revised, principled top level taxonomy is proposed [4]. We analyzed the representation of health disorders in WordNet [5, 6]. Many concepts representing health disorders in medical terminologies, when present in WordNet, are also categorized as health disorders in WordNet (e.g., *Leukemia* is a hyponym of *Cancer*). However, we noted that some medical signs or symptoms are just hyponyms of a generic concept representing the mechanism involved in the abnormal phenomenon. For example, the hypernym of *Bronchospasm* in WordNet is *Constriction*. The emphasis is put on the general physical mechanism involved in the bronchospasm rather than on its pathological aspects. As a consequence, there is no relationship between *Bronchospasm* and the biomedical domain in WordNet.

III. Examples of medical ontologies

A. GALEN

Introduction

GALEN (Generalised Architecture for Languages, Encyclopaedias, and Nomenclatures in medicine) is a European Union project that was initiated in the early nineties and seeks to provide re-usable terminology resources (in a broad sense) for clinical systems. A key feature of the GALEN approach is that it is a top-down one, first defining knowledge top levels and the representation formalism, then populating the ontology. GALEN relies on an ontology, the Common Reference Model, formulated in a specialized description logic, the GALEN Representation and Integration Language (GRAIL). An important additional focus of GALEN has been in developing tools and techniques allowing the information found in existing coding and classification schemes to be mapped to the GALEN Common Reference Model and in encompassing linguistic resources representing several languages. The model aims to represent "all and only sensible medical concepts". GALEN is not a repository of every kind of information used in the practice of medicine, rather an attempt to represent the underlying conceptual model of medicine, independently of any specific application. OpenGALEN provides a point of access to the GALEN Common Reference Model and to descriptions and specifications of the GALEN technology. While the scope of biomedicine has not been entirely covered yet, the coverage of surgical procedures is extensive in the current release (October 2001).

Ontological features

The GALEN concept model relies on a hierarchy of elementary categories that serve for complex descriptions, e.g., descriptions of medical procedures [7]. The top level categories are presented in Figure 3. The major subdivision is between *Phenomenon*, subsuming structures, processes and substances, and *ModifierConcept*. The notion of *ModifierConcept* is used to distinguish from concepts which represent things that can exist on their own (e.g., physical objects, ideas) the concepts that only make sense when linked to another object as modifiers (e.g., severe), modalities (e.g., presence, uncertainty) or collections (e.g., polyps, as opposed to polyp). GALEN partitive relations involving physical parts are of several kinds [8], including *has surface division*, (e.g., `Hand has-surface-division Palm`), *has solid division* (e.g., `Heart has-solid-division Cardiac Septum`), *has layer* (e.g., `Stomach has-layer Mucosa`), *has blind pouch division* (e.g., `Caecum has-blind-pouch-division`

Appendix), *has linear division* (e.g., `Intestine has-linear-division Jejunum`), *has structural component* (e.g., `Knee Joint has-structural-component Meniscus`), *is made of* (e.g., `Meniscus is-made-of Fibrocartilage`). Furthermore, another principle in GALEN is compositionality. For example, rather than representing *Pole of kidney* as a category, it is represented as a defined composite, i.e., `Pole which is-structural-component-of Kidney`, since there seemed to be sufficient commonality in a notion such as pole to merit capturing it individually.

B. Unified Medical Language System

Introduction

The Unified Medical Language System® (UMLS®), has been developed and maintained by the U. S. National Library of Medicine since 1990. It is intended to help health professionals and researchers use biomedical information from different sources [9]. The UMLS comprises two major inter-related components: the Metathesaurus®, a huge repository of concepts, and the Semantic Network, a limited network of semantic types. The current version (2002AA) of the Metathesaurus integrates about 775,000 concepts from more than sixty families of vocabularies [10]. While the structure of each source vocabulary is preserved, terms that are equivalent in meaning are clustered into a unique concept. Interconcept relationships are either inherited from the source vocabularies or specifically generated. The UMLS building process imposes no restrictions on the source vocabularies prior to integrating their terms and structure into the Metathesaurus. As a result, the UMLS Metathesaurus is not an ontology, since it does not provide the level of organization of concepts that is expected from ontologies. In contrast, the UMLS Semantic Network is a high-level representation of the biomedical domain based on Semantic Types (STs) under which all the Metathesaurus concepts are categorized, and which is intended to provide a basic ontology for the biomedical domain [11] (Figure 4).

Ontological features (Semantic Network)

The UMLS Semantic Network is a network of 134 semantic types used to categorize Metathesaurus concepts. Each UMLS Semantic Type has a definition. The semantic types are organized in two single-inheritance hierarchies, one for entities, one for events. The *is-a* link allows nodes to inherit properties from higher-level nodes. In addition, associative relationships are instantiated between the semantic types. They represent general high-level, definitional knowledge, such as “drugs treat diseases”. Relationships between Semantic Types define the allowable semantics for relationships between Metathesaurus concepts [12]. For example, a particular drug may treat a particular disease. Besides the taxonomic relation, associative relations are divided into five subcategories: Physical (e.g., *part_of*, *branch_of*, *ingredient_of*), Spatial (e.g., *location_of*, *adjacent_to*), Functional (e.g., *treats*, *complicates*, *causes*), Temporal (*co-occurs_with*, *precedes*), and Conceptual (e.g., *evaluation_of*, *diagnoses*). At the highest level, the UMLS Semantic Network is built on the opposition of entities and events; the next level distinguishes between *Physical object* and *Conceptual entity* as entities, and *Activity* and *Phenomenon or process* as events (Figure 5).

Each Metathesaurus concept is assigned one or more Semantic Type(s). The economy principle, which is close to the principle of parsimony [13] and Swartout's principles [14], has been applied, resulting in three rules:

R1. *Assign the most specific semantic type available*. The level of granularity varies across the UMLS Semantic Network. The intent is to establish a set of semantic types, which are useful for a variety of tasks without introducing undue complexity. The most specific semantic type in the semantic type hierarchy is assigned to the concept [15].

R2. *Assign multiple semantic types if necessary*. Instead of creating a lattice structure, with hybrid types inheriting from two supertypes, the Semantic Network has a single inheritance tree structure. As a consequence, a Metathesaurus concept inheriting from two semantic types is assigned to both types.

R3. *Assign a less specific semantic type (supertype) if no more specific semantic type (subtype) is available*. Rather than proliferating the number of semantic types to encompass additional subcategories, concepts that cannot be categorized by any sibling semantic type are simply assigned their common supertype [11, 16].

The consequences of applying the economy principle on the representation of knowledge in the UMLS were presented in a previous paper [17].

C. Systematized Nomenclature of Medicine

Introduction

Following the principles implemented in the Systematized Nomenclature of Pathology®, whose multiaxial model made of four axes (i.e., topography, morphology, etiology, function) was created in 1965, the Systematized Nomenclature of Medicine® (SNOMED®) is an inventory of medical terms and concepts that has been designed by the College of American Pathologists[18]. The first version of SNOMED dates back to 1979. Terms in SNOMED are detailed and assigned to eleven independent modules (fields), each of which is systematized. SNOMED® Reference

Terminology (RT) has been designed to complement the coverage of medical concepts in SNOMED with additional features including multiple hierarchies and semantic definitions[19]. Its objective is to allow the full integration of all medical information in the electronic medical record into a single data structure, facilitating interoperability between a wide variety of systems and clinical records. SNOMED RT combines the granularity and comprehensiveness of SNOMED terms and term codes with formal features. SNOMED RT currently consists of a set of 121,000 concepts and 340,000 relationships. The College of American Pathologists has entered into a collaboration with the United Kingdom's National Health Service Executive (NHS) to combine SNOMED RT and Clinical Terms Version 3 of the NHS thesaurus of health care terms (also known as Read Codes V3). The new work is named SNOMED Clinical Terms (abbreviated SNOMED CT).

Ontological features

SNOMED RT relies on a multiaxial model. Each concept is given a semantic definition stated in description logic [20]. Similarly, SNOMED CT represents multiple hierarchies (Figure 6), e.g., *Candidal meningitis* is a child of both *Central nervous system candidiasis* and *Fungal meningitis*. Additional knowledge is provided by associative relationships, which are inspired by the former SNOMED multiaxial descriptions. For example, the concept *Candidal meningitis* is related to several other SNOMED concepts that play distinct roles:

Causative agent	Candida (organism)
Pathological process	Infectious disease (disorder)
Associated morphology	Inflammation (morphologic abnormality)
Finding site	Meninges structure (body structure)

For each kind of concept, a pattern gives the authorized roles, i.e., the semantic relations allowed to be instanciated. For example, *associated morphology*, *causative agent, pathological process*, *finding site* and *is-a*, are some of the roles that are allowed for *Disease* category. The same approach allows the representation of *part of* hierarchies, e.g., *part of* is a role for *Anatomical structure* category.

D. Digital Anatomist Foundational Model

Introduction

The Digital Anatomist is an ontology of anatomy that has been in development since 1997 at the University of Washington [21]. It was initiated as an enhancement of the anatomical content of the UMLS Semantic Network and

Metathesaurus, and has subsequently evolved into the Digital Anatomist Foundational Model [22]. The component of the model currently distributed as part of the UMLS is known as the University of Washington Digital Anatomist (UWDA). The objective is to provide a conceptualization of the material objects and spaces that constitute the human body; the representation should be parsable by machine, and also be comprehensible by both expert and novice users of anatomical information. Anatomy being quite pervasive in medicine, such an ontology could benefit knowledge representation in virtually all biomedical subdomains. The Digital Anatomist Anatomical ontology (Ao) is made of nearly 60,000 concepts. Originally limited to gross anatomy, the Foundational Model is now being extended to the cellular and subcellular levels.

Ontological Features

Definitions of physical anatomical entities were formulated by specifying constraints [23] in terms of their spatial dimension, mass, and inherent 3D shape, as well as the structural units that make up the body. Relationships in Digital Anatomist Foundational Model are constrained to those that represent the structural organization of physical anatomical entities. The top level of Digital Anatomist is *Anatomical entity*, which is divided into *Physical anatomical entity* and *Conceptual anatomical entity*. Conceptual entities, which do not have spatial dimension, include a taxonomy of anatomical relationships and concepts such as *Developmental stage* and *Muscle action*. The first level of the Digital Anatomist taxonomy of physical anatomical entities, which have spatial dimension, is given in Figure 7.

A distinction is made between the physical entities that have mass, such as anatomical structures and body substances (*Material physical anatomical entity*), and those that do not have mass (i.e., anatomical spaces, surfaces, lines and points, classified as *Non-material physical anatomical entity*). The attribute of inherent 3D shape contrasts anatomical structures – which are objects – with body substances. While anatomical structures have an inherent shape, body substances assume the shape of the anatomical structures that contain them. The Digital Anatomist Foundational Model integrates the Anatomical ontology (Ao) with two much smaller ontologies: the Physical state ontology and the Spatial ontology (So). The latter represents geometric objects and 3D shape classes, and also distinguishes between bona fide (real) and fiat (virtual) boundaries of volumes, surfaces and lines [24].

In addition to taxonomic relationships, anatomy requires meronomies and a number of other relationships that describe the spatial organization of anatomical structures and substances. In Digital Anatomist, partitive hierarchies have been formulated, using the transitive *part-of* relation. In addition, two anatomical relations, *branch-of* and *tributary-of*, have been defined to represent relationships among tree-like structures such as nerves, arteries, veins, and lymphatic vessels. The Foundational Model extends these relationships to boundary, orientation, connectivity and location; the latter specified as containment, adjacency and anatomical coordinates [25].

E. MENELAS ontology

Introduction

MENELAS is a European Union project that was developed between 1992 and 1995. It was conceived as an access system for medical records using natural language in several European languages [26]. MENELAS adopted a knowledge-base approach to natural language understanding, and relies on a body of knowledge expressed in Conceptual Graphs. The test domain for the project was coronary artery diseases. Resources developed as part of the MENELAS project include linguistic knowledge bases (domain-specific syntactic and semantic lexicons for several languages), and medical knowledge bases (ontology of the domain of coronary artery diseases, encyclopaedic knowledge attached to each concept in the form of schemata, etc). The ontology, made of 1800 types and 300 relations, was acquired from several sources: interviews with physicians, reuse of existing terminological resources, and corpus analysis.

Ontological Features

From a structural viewpoint, the MENELAS ontology was initially developed as a lattice, according to the Conceptual Graph model [27]. To avoid the ambiguities due to multiple inheritance in a lattice, the principles of opposition of siblings and unique semantic axis have later been adopted, leading to a tree structure [28]. In doing that, the authors of the MENELAS ontology intended to organize all the concepts according to their genus and differentia. The top level of the MENELAS ontology is shown in Figure 8. Concept labels in MENELAS are simply mnemonic tags created for convenience rather than to convey meanings. The actual meaning of a concept comes from its position in the hierarchy, and can be interpreted by reading its documentation. For example, in MENELAS, *Physical object* is an *Abstract object*, whereas most ontologies oppose the two concepts. In fact, *Abstract object* is defined as a *Substratum* that has instances in the world, and it is opposed to *Ideal object*. For example, *Apple* is an *Abstract object* whereas *Two* is an *Ideal object*. Relations are categorized according to the kind of concepts they can link. Relations between physical objects include, for example, relations between mass objects and countable objects (*contains*, *has for dosage*, and *constituted of*), and relations from real objects to pseudo-objects (*component of*). The *part of* relation may link any kinds of physical objects and may specialize into *part fragment*, and *part segment*. Another relation, *functional part*, allows functional viewpoints to be represented. Models and schemas provide additional knowledge. However, the validity of this additional knowledge may be limited to the context of the MENELAS application, i.e., domain-specific and task-oriented. For example, the model for organ component includes the notion of duct, e.g., for blood vessels.

IV. Representations of the concept *Blood*

Having presented the characteristics and top level organization of several biomedical ontologies, we now examine how one particular concept is represented in the different systems. This concept must be present in all systems and intuitively illustrative of the issues faced when comparing several representations. We selected the concept *Blood* for these reasons. The comparison of the representations is based on the definitions of the concept (textual and formal), as well as its properties.

Definitions of *Blood* provided by medical dictionaries include:

- The fluid that circulates through the heart, arteries, capillaries, and veins, carrying nutriment and oxygen to the body cells (Dorland's)
- The "circulating tissue" of the body; the fluid and its suspended formed elements that are circulating through the heart, arteries, capillaries, and veins (Steadman's).

Steadman's definition foreshadows the duality of representation of *Blood* as both a tissue and a fluid.

A. Representation in OpenCyc

Although not represented as a type in the OpenCyc ontology, *Blood* is given as an example of subordinates of *Mixture*, as well as mud, air and carbonate beverage (another meaning of *Blood*, referring to the notion of lineage, is represented as such in OpenCyc). Every *Mixture* is a *Tangible stuff* composed of two or more different constituents which have been mixed. These constituents do not form chemical bonds, and later the mixture may be resolved by some separation event. In other words, a mixture has a composition but no structure. As a mixture, *Blood* is an element of the collection *Existing stuff Type* (`#$isa #$Mixture #$ExistingStuffType`), which means that blood is temporally and spatially stuff-like. This implies a fundamental feature: division in time or space does not destroy the stuff-like quality of the object. Another example of stuff-like biological thing given by OpenCyc authors is *Striated muscle*. *Mixture* is a *Tangible thing* (`#$genls #$Mixture #$TangibleThing`), i.e., a thing made of some sort of matter and whose nature is primarily material (in the sense that it does not have important non-physical properties, such as encoded information). In OpenCyc, *Blood* is represented differently from concepts such as *Sweat* and *Semen*, which are subordinates of *Bodily secretion*. In addition, *Sweat* is also a subordinate of *Excretion substance*, which means it is considered a waste.

B. Representation in WordNet

In WordNet, *Blood* has the following definition: "the fluid (red in vertebrates) that is pumped by the heart. Blood carries oxygen and nutrients to the tissues and carries waste products away; the ancients believed that blood was the seat of the emotions". Five other meanings of the word blood are also represented, e.g., temperament or disposition. The direct hypernym of *Blood* is *Liquid body substance*, The list of all hypernyms of *Blood* is given in Figure 9 (a). In WordNet, there is no significant difference in the categorization of *Blood*, *Sweat*, and *Semen*. All of them are categorized as *Liquid body substance*. Unlike *Blood*, *Sweat* is linked to *Liquid body substance* through the synset *Secretion*.

C. Representation in OpenGALEN

In OpenGALEN, *Blood* is a subordinate of *Soft tissue* as well as *Lymphoid tissue*, *Integument*, and *Erectile tissue* among others. The hierarchical environment of *Blood* in GALEN is given in Figure 9 (b). Remarkably, this hierarchical structure is actually a lattice since *Substance* is the common subtype of *Generalised substance* and *Substance or physical structure*, both subtypes of *Phenomenon*. In GALEN, *Blood* is represented differently from *Sweat* and *Semen*, which are subordinates of *Body substance*.

D. Representation in the UMLS

In the UMLS, *Blood* is found as a concept in the Metathesaurus. It is assigned the Semantic Type *Tissue*, defined as "An aggregation of similarly specialized cells and the associated intercellular substance. Tissues are relatively non-localized in comparison to body parts, organs or organ components". In the Semantic Network, *Tissue* is a subordinate of *Fully Formed Anatomical Structure*. The whole *is-a* hierarchy for *Blood* is given in Figure 9 (c). In the UMLS, *Blood* is not assigned the same Semantic Type as concepts such as *Sweat* and *Semen*, which are categorized as *Body Substance*. Moreover, in the UMLS Metathesaurus, ancestors of *Blood* include *Body fluid*, *Body substance*, *Soft tissue* and *Connective tissue*.

E. Representation in SNOMED CT

In SNOMED CT, *Blood* is found in the concept category *Substance* as a subordinate of *Blood material*, as well as *Blood component*. The hierarchical environment of *Blood* in SNOMED CT is given in Figure 9 (d). Multiple inheritance in SNOMED CT allows *Body fluid*, an ancestor of *Blood*, to inherit from both *Body substance* and *Liquid substance*. These two concepts are descendants of the top level category *Substance*. Subordinates of *Body fluid* also include *Sweat*, *Lymph*, and *Platelet rich plasma*.

F. Representation in the Digital Anatomist Foundational Model

In Digital Anatomist, *Blood* is a subordinate of *Body substance*, defined as "a material physical anatomical entity in a gaseous, liquid, semisolid or solid state, with or without the admixture of cells and biological macromolecules; produced by anatomical structures or derived from inhaled and ingested substances that have been modified by anatomical structures as they pass through the body". In their representation of anatomy, the authors of Digital Anatomist constrain modeling to a strictly structural context. They declare *Organ* as the organizational unit for macroscopic anatomy, and define it as "an anatomical structure that consists of the maximal set of organ parts so connected to one another that together they constitute a self-contained unit of macroscopic anatomy, distinct both morphologically and functionally from other such units". Organs have mass and an inherent 3D shape. Other direct subordinates of *Anatomical structure* either constitute organs (and are classified as *Organ part*) or are constituted by organs (and are classified as *Body part*, such as head and limb, and *Organs system*, such as respiratory or skeletal system). In addition, *Cell* and *Biological macromolecule*, which can be a part of body substances as well as of anatomical structures, are also direct subordinates of *Anatomical structure*. The smallest organ part is *Tissue*, which, along with all the other subordinates of *Anatomical structure*, inherits <u>inherent 3D shape</u> as a property from its parent. Thus, because body substances lack the defining attribute <u>inherent 3D shape</u>, body substances, including *Blood*, form a lineage of the Digital Anatomist taxonomy that is completely distinct from the lineage containing anatomical structures, such as organs and organ parts, including tissues. The whole *is-a* hierarchy for *Blood* is represented in represented in Figure 9 (e). According to its definition in Digital Anatomist, in addition to *Blood*, *Body substance* also includes other cellular fluids, such as *Semen*, as well as *Secretions* (e.g., saliva and sweat), *Transudates* (e.g., tissue fluid, lymph and cerebrospinal fluid), *Excretions* (e.g., feces, urine), along with such substances as inhaled air, intercellular matrix, and cell substance (e.g., cytosol). All these body substances are defined in terms of the anatomical structures that process and contain them.

G. Representation in MENELAS

In MENELAS, *Blood* is a subordinate of *Body fluid*. The only other body fluid represented in MENELAS is *Lymph*. The whole ascendance of *Blood* in MENELAS is represented in Figure 9 (f). The concept *Mass object,* one of the ancestors of *Blood*, has three subtypes: *Agglomerat* (divided into *Inorganic agglomerat* and *Organic agglomerat*), *Substance* (divided into *Biochemical substance* and *Chemical substance*), and *Tissue* (divided into *Body fluid* and *Connective tissue*). Thus *Blood* belongs to a branch that is different from the branch containing *Substance*. Furthermore, *Tissue*, defined as a set of cells, is opposed to *Substance*, defined as a set of molecules. A model, i.e., additional knowledge, is associated with the concept *Body fluid*. This model emphasizes a given property of fluids, namely <u>viscosity</u>, that was of interest in the context of the MENELAS application. The representation provided by MENELAS is ad hoc, e.g., the distinction among concepts is based on properties (for example, <u>viscosity</u>) that are useful for problem solving and natural language understanding. Types have context-dependent definitions and characteristics, e.g., *Body fluid* is considered a *Tissue* in MENELAS in a very unusual way, since other ontologies separate fluids and substances from tissues. Not surprisingly, in this application ontology for the interpretation of coronary angiography reports, *Semen* is out of the scope of MENELAS. *Sweat* is not categorized as *Substance* but as *Cutaneous sign* (i.e., sweating).

V. Discussion

This study has revealed some ontological issues dealing with compatibility among systems. These issues are examined in the particular context of the biomedical domain whose characteristics may have an influence on ontology design. The additional knowledge represented in some ontologies will be discussed as well. Compatibility issues will lead us to compare two approaches to building ontologies: unifying representations starting from existing ontologies and creating ontologies from a theory of what exists in the domain to be represented.

A. Compatibility among representations

The differing representations of *Blood* in several systems raise issues about compatibility among ontologies. Obviously, the representation of most concepts is simpler than that of *Blood*, and different ontologies often provide roughly similar views. For example, *Heart,* another concept central to the biomedical domain, is consistently represented as some subordinate of *Organ*. What makes the representation of *Blood* more complex is that two different superordinates are found: *Tissue* and *Body substance*. GALEN and the UMLS Semantic Network categorize *Blood* as *Tissue* while the Digital Anatomist Foundational Model categorizes it as *Body substance*. In between, WordNet, SNOMED and MENELAS categorize *Blood* as *Body fluid*, itself categorized as *Body substance* in WordNet and SNOMED, but as *Tissue* in MENELAS. Finally, in GALEN, *Tissue* and *Body substance* are two subtypes of *Substance*. A composite representation of *Blood* is shown in Figure 10.

Superficially, this dual representation of *Blood*, as both *Tissue* and *Body substance*, does not reveal any major incompatibility, such as, for example, circular hierarchical relationships. However, a representation in which *Blood* is a common subtype of *Tissue* and *Body substance* would not fit the additional structural constraint of opposition of siblings. Analyzed more carefully, the definitions for *Tissue* in the Digital Anatomist Foundational Model and the Semantic Network, although closely related, are not equivalent. In Digital Anatomist, *Tissue* is defined as “an organ part that consists of similarly specialized cells and intercellular matrix, aggregated according to specific spatial relationships; together with other tissues, it constitutes an organ component”. This definition is largely based on that of UMLS (“An aggregation of similarly specialized cells and the associated intercellular substance…”), but Digital Anatomist further constrains the concept of *Tissue* by specifying a requirement for a specific spatial organization of the cells that constitute a given kind of tissue. Various types of epithelium, muscle tissue and neural tissue fulfill this requirement. This additional criterion is particularly relevant for disambiguating the classification of *Blood* in Digital Anatomist. Broadly speaking, blood cells are similarly specialized, but if they are allowed to aggregate in anticoagulated blood, as they do when they settle or are centrifuged, they do not constitute a tissue in terms of the Digital Anatomist Foundational Model definition. Yet, they would form a tissue in terms of the UMLS definition. *Blood* is classified as a *Tissue* not only by GALEN and UMLS, but also by textbooks of histology. The difficulty with such a classification is illustrated by the fact that none of these sources classify lymph, cerebrospinal fluid or semen as tissues. Yet, their essential nature (i.e., the suspension of similarly specialized cells in a body substance of liquid state) corresponds with that of blood. While *Tissue* is a child of *Organ part* in Digital Anatomist, it is a sibling of *Body part, organ or organ component* in the UMLS Semantic Network whose definition of *Tissue* states that “…Tissues are relatively non-localized in comparison to body parts, organs, or organ components”. With a lesser coverage of the biomedical domain, OpenCyc hooks *Blood* directly to *Mixture*. Although specialized types of tissues such as *Muscle tissue* and *Fatty tissue* are defined, there is no generic *Tissue* type in OpenCyc. Such discrepancies, small as they are, make the alignment difficult and can lead to possibly conflicting representations.

Ad hoc representations are often present in application ontologies. For example, in MENELAS, *Real object* is a *Physical object*, itself being a kind of *Abstract object*. These assertions apparently contradict the general axioms present in other ontologies such as “no abstract object has a location in space”, “no abstract object occurs at a point in time”, and “physical occurs in time and space”. As mentioned earlier, in MENELAS, *Abstract object* is opposed to *Ideal object*, while, in many ontologies, it is opposed to *Physical object*. As a consequence, *Abstract object* in MENELAS has a specific meaning and cannot be simply aligned with the type that bears the same name in other existing ontologies. Local definitions and an ad hoc organization are usually not detrimental to application ontologies whose major goal is to support problem solving in a specific context rather than to represent knowledge independently and consistently. While they may be useful for application ontologies, these features are likely to cause problems if these ontologies need to be shared by different applications, or linked to other ontologies.

B. Additional knowledge

In our example, OpenCyc, GALEN, SNOMED CT, the UMLS and MENELAS all provide additional general knowledge related to *Blood*. As a *Mixture* in OpenCyc, *Blood* belongs to the set of stuff-like things. It is made of constituents, and the type of events categorized as *Separation mixture* can apply to it. This reflects the fact that, in some situation, blood constituents can separate out spontaneously. Erythrocyte sedimentation, whose rate has been used as an indicator of inflammatory processes, is an example of spontaneous, reversible separation of blood components. In SNOMED CT, *Blood* is involved in the description of several other entities by the means of specific roles. For example, *Blood* is a *Finding site* for *Bacteremia*. As *Body fluid* in MENELAS, *Blood* acquires the viscosity property. At an upper level, *Blood* is also a subtype of *Mass object*. As a *Mass object*, it inherits general knowledge that is represented for this type (e.g., *Mass object* may be a component of *Countable object*). From *Mass object*, *Blood* also inherits the quantity property that can be expressed with quantitative values and units. A specific feature in GALEN is to identify two distinct physical states for *Blood*, *Liquid blood* and *Coagulated blood*, represented as descendants of *Blood*. In addition, *Blood* inherits the properties of *Tissue* (e.g., `irradiation acts on tissue`) and Substance (e.g., `substance has a mass`). Additionally, GALEN extends the representation of *Blood* through roles such as *hasCountability infinitelyDivisible*, meaning that *Blood* is not a discrete object. By categorizing *Blood* as *Tissue* in the UMLS, potential relationships with other kinds of entities can be inferred from the Semantic Network. Relationships of *Tissue* to other Semantic Types, result in predicates including `Tissue produces Biologically Active Substance`, `Tissue produces Body Substance`, `Tissue is a` `location of Pathologic Function`, `Embryonic Structure is a developmental form of Tissue`, and `Tissue surrounds Tissue`.

More than to a particular kind of ontology, the presence of general knowledge seems to be related not only to its intent to support knowledge processing, but also to its integration into a set of resources available to achieve this goal. For example, the notion of <u>countability</u> represented in MENELAS and GALEN ontologies (i.e., discrete objects or not) does not appear in the UMLS Semantic Network. The SPECIALIST Lexicon, another knowledge source in the UMLS, provides this property for each of its entries, making it available for computing inflectional variants (e.g., plural form). However, properties that hold on concepts are different from properties that hold on English words. For example, *Onion* and *Garlic* are concepts that have similar properties while the word “onion” is countable and the word “garlic” is uncountable.

Finally, even when taxonomies look compatible at first sight, additional properties may be inconsistent or conflicting among ontologies. For example, in GALEN, <u>liquid</u> is true only for *Liquid blood* and not for *Blood* itself, while in SNOMED, *Blood* inherits the <u>liquid</u> property from its supertype *Liquid substance*.

C. Unification of existing ontologies vs. theory of reality

As illustrated in this paper, several major efforts have been made in the last fifteen years to produce ontologies, including biomedical ontologies. The resulting systems provide as many partitions of the biomedical domain, i.e., different cuts through the same reality. Different images of a specimen can be obtained with a microscope when modifying the magnification or applying filters. Likewise, choices made regarding granularity or the selection of features to be represented may result in different ontologies. However, observing the same reality should result in producing compatible, if different, ontologies. Conversely, ontologies developed for solving problems in a given application may provide a conceptualization valid only in the context of this application, but not sharable.

As more diverse systems are developed and more groups are involved in sharing medical information, one solution would be to put those sets of concepts, relations and framework together within a single system. A common ontology that would be shared in common by a plurality of systems could rise from merging diverse conceptualizations. However, while merging conceptualizations in a given domain, two situations have to be considered: merging different representations of a single theory of the domain, and merging different theories of the domain. Several technical issues arise from attempting to merge different representations of a domain, even when representations occur within a single theory of the domain. While difference in granularity is usually not a problem, differing naming conventions and the lack of reliable textual definitions may result in merging difficulties. Tools, e.g., SMART[29], have been developed to assist the ontology developer in merging existing ontologies.

Merging ontologies that convey different theories of the domain requires that the target system be able to represent and clearly identify microtheories. Differences in domain theory prevent the intermediate levels of ontologies from being compatible. By the means of microtheories, assertions and hierarchies are given a context. Therefore, different theories can be represented without confusion among them. For example, in WordNet 1.6, the direct hypernym of *Blood* was *Humor*, which refers to a pre-scientific representation of the human body. Humors are the four fluids in the body whose balance was believed to determine our emotional and physical state (the humors are blood, phlegm, and yellow and black bile). While knowledge such as `Blood is a Humor` is useful to understand medieval texts, it is of no interest in a scientific biomedical ontology. Rather than clustering *Humor* and *Liquid body substance* into a single concept, typically, "humors" must be identified as part of a microtheory. Similarly, specific views are needed to represent lay knowledge, or to encompass oriental medicine that provides representation of *Blood* as a fluid manifestation of Qi.

Another point is that inconsistencies can be reduced by using philosophical analyses of notions such as identity or unity[30]. The tools of formal ontology can help solve practical problems in medical applications. For example, mereotopology, the theory of parts and boundaries, already applied to geographic information systems, could also help represent anatomical structures and subdivisions of the human body. Besides these general theories, several features of the biomedical domain, such as the abundance of empirical concepts, make it an ontological challenge to develop a theory of the biomedical domain. A formal approach would certainly result in clarifying some key concepts, such as *Tissue*, that are broadly used to sustain classifications in biomedicine.

We believe that these two approaches, i.e., merging existing ontologies and developing an ontology from the ground up, are complementary rather than conflicting solutions. Illustrative of each approach, we would like to mention two ongoing projects. Taking advantage of the UMLS, the Medical Ontology Research project[†] at the U.S. National Library of Medicine analyzes the set of concepts that are shared among most of the source terminological systems. It is expected to provide a better understanding of the core concepts in the biomedical domain. On the other

[†] http://lhncbc.nlm.nih.gov/cgsb/research/umls/mor/

hand, the work of IFOMIS[‡] at the University of Leipzig, Germany consists of building ontological theories in the field of clinical trials, relying on a formal approach.

VI. Conclusion

Although general ontologies and limited application ontologies may be useful in biomedicine, what many applications in biomedicine (from natural language processing to information retrieval) would most benefit from is a domain ontology, i.e., a sound, large-coverage ontology of the biomedical domain. We examined the biomedical ontologies currently available and found that none of them fully meets these requirements. Moreover, we observed a certain lack of compatibility among their representations. Several factors contribute to this situation. First, there is no agreement on an upper level ontology to which a biomedical ontology could hook its concepts. Second, there is no unique theory of the domain, and some characteristics of biomedicine make it particularly difficult to represent (e.g., large number of concepts and inherent vagueness of some concepts). Finally, pragmatic aspects rather than formal principles prevail in the design on some biomedical ontologies. The road to the biomedical ontology of our dreams may well be long and paved with difficulties. Meanwhile, we believe that identifying and clarifying the core concepts of the domain will contribute to improve the sharability of existing ontologies as well as the interoperability of the applications that rely on them.

VII. Acknowledgments

This research was supported in part by an appointment to the National Library of Medicine Research Participation Program administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the National Library of Medicine.

The authors would like to thank Jeremy Rogers, Cornelius Rosse, Jacques Bouaud and Pierre Zweigenbaum for providing valuable insights about the ontologies they contributed to create. Thanks to Tom Rindflesch for his encouragement and advice.

VIII. References

1. Burgun A, Bodenreider O. Ontologies in the biomedical domain. Part I: Principles. *JAMIA.* 2002;J Am Med Inform Assoc:(*** to be completed ***).

2. Bodenreider O. *Medical Ontology Research (Report to the Board of Scientific Counselors).* Bethesda, Maryland: Lister Hill National Center for Biomedical Communications; 17 May 2001 2001.

3. Fellbaum C, ed. *WordNet: An electronic lexical database*. Cambridge, Massachusets: MIT Press; 1999.

[‡] http://ifomis.de/

4. Gangemi A, Oltramari A. A formal ontology approach to refine lexical taxonomies: the case of WordNet top level. In: Welty C, Smith B, eds. *Collected papers from the Second International Conference "Formal Ontology in Information Systems"*: ACM Press; 2001:285-296.

5. Burgun A, Bodenreider O. Comparing terms, concepts and semantic classes in WordNet and the Unified Medical Language System. *Proc NAACL Workshop, "WordNet and Other Lexical Resources: Applications, Extensions and Customizations".* 2001:77-82.

6. Burgun A, Bodenreider O. Mapping the UMLS Semantic Network into General Ontologies. *Proc AMIA Symp.* 2001:81-85.

7. Trombert-Paviot B, Rodrigues JM, Rogers JE, et al. GALEN: a third generation terminology tool to support a multipurpose national coding system for surgical procedures. *Int J Med Inf.* 2000;58-59:71-85.

8. Rogers J, Rector A. GALEN's model of parts and wholes: experience and comparisons. *Proc AMIA Symp.* 2000:714-718.

9. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med.* 1993;32(4):281-291.

10. UMLS. *UMLS Knowledge Sources*. 12th ed. Bethesda (MD): National Library of Medicine; 2001.

11. McCray AT, Nelson SJ. The representation of meaning in the UMLS. *Methods Inf Med.* 1995;34(1-2):193-201.

12. McCray AT, Bodenreider O. A conceptual framework for the biomedical domain. In: Sung M, Green R, eds. *Semantics of Relationships*: Kluwer; 2001:(to appear).

13. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies.* 1995;43(5-6):907-928.

14. Swartout B, Patil R, Knight K, Russ T. Toward Distributed Use of Large-Scale Ontologies. Paper presented at: Proceedings of the 10th Workshop on Knowledge Acquisition, Modeling and Management, 1996; Banff, Canada.

15. McCray AT, Hole WT. The scope and structure of the first version of the UMLS Semantic Network. *Proc Annu Symp Comput Appl Med Care.* 1990:126-130.

16. McCray AT. Representing biomedical knowledge in the UMLS Semantic Network. In: Broering NC, ed. *High-performance medical libraries*: Meckler; 1993:45-55.

17. Burgun A, Bodenreider O. Aspects of the taxonomic relation in the biomedical domain. In: Welty C, Smith B, eds. *Collected papers from the Second International Conference "Formal Ontology in Information Systems"*: ACM Press; 2001:222-233.

18. Rothwell DJ, Cote RA. Managing information with SNOMED: understanding the model. *Proc AMIA Annu Fall Symp.* 1996:80-83.

19. Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care. *Proc AMIA Annu Fall Symp.* 1997:640-644.

20. Spackman KA. Normal Forms for Description Logic Expressions of Clinical Concepts in SNOMED RT. *Proc AMIA Symp.* 2001:627-631.

21. Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and organizational principles for anatomical knowledge representation: the digital anatomist symbolic knowledge base. *J Am Med Inform Assoc.* 1998;5(1):17-40.

22. Rosse C, Shapiro LG, Brinkley JF. The digital anatomist foundational model: principles for defining and structuring its concept domain. *Proc AMIA Symp.* 1998:820-824.

23. Michael J, Mejino JL, Jr., Rosse C. The role of definitions in biomedical concept representation. *Proc AMIA Symp.* 2001:463-468.

24. Mejino JL, Jr., Rosse C. Conceptualization of anatomical spatial entities in the Digital Anatomist Foundational Model. *Proc AMIA Symp.* 1999:112-116.

25. Mejino JL, Jr., Noy NF, Musen MA, Brinkley JF, Rosse C. Representation of structural relationships in the Foundational Model of anatomy. *Proc AMIA Symp.* 2001:973.

26. Zweigenbaum P. Menelas - an Access System for Medical Records Using Natural- Language. *Computer Methods and Programs in Biomedicine.* 1994;45(1-2):117-120.

27. Bouaud J, Bachimont B, Charlet J, Zweigenbaum P. Acquisition and structuring of an ontology within conceptual graphs. Paper presented at: ICCS'94 Workshop on Knowledge Acquisition using Conceptual Graph Theory, 1994; University of Maryland, College Park, MD.

28. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF. Issues in the Structuring and Acquisition of an Ontology for Medical Language Understanding. *Methods of Information in Medicine.* 1995;34(1-2):15-24.

29. Noy NF, Musen MA. An algorithm for merging and aligning ontologies: automation and tool support. Paper presented at: Proceedings of the 16th National Conference on Artificial Intelligence (AAAI-99) Workshop on Ontology Management, 1999; Orlando, Florida.

30. Guarino N, Welty C. A Formal Ontology of Properties. Paper presented at: EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management, 2000.

IX. Appendix

References for ontologies and terminologies mentioned in this paper

(ULRs valid as of June 12, 2002)

Digital Anatomist	http://sig.biostr.washington.edu/projects/da/
MENELAS	http://www.biomath.jussieu.fr/~pz/Menelas/
OpenCyc™	http://www.opencyc.com/
OpenGALEN	http://www.opengalen.org/
SNOMED® CT	http://www.snomed.org/
UMLS®	http://umlsks.nlm.nih.gov/ (free UMLS registration needed)
WordNet®	http://www.cogsci.princeton.edu/~wn/

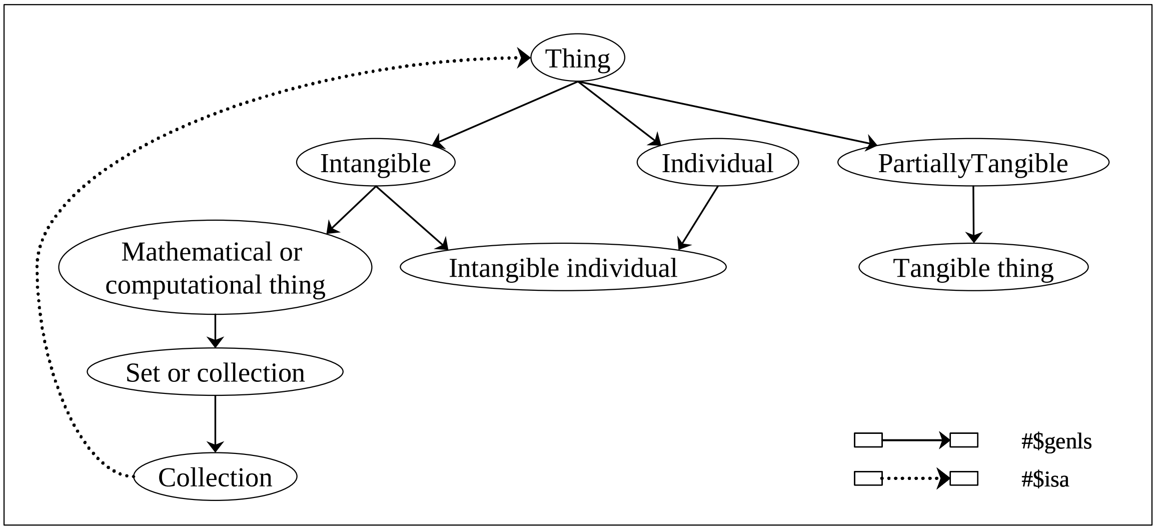

Figure 1 - Top level in OpenCyc (partial representation)

Abstraction
Activity
Entity
Event
Group
Location
Natural phenomenon
Possession
Psychological feature
Shape
State

Figure 2 - Top level in WordNet ("unique beginners")

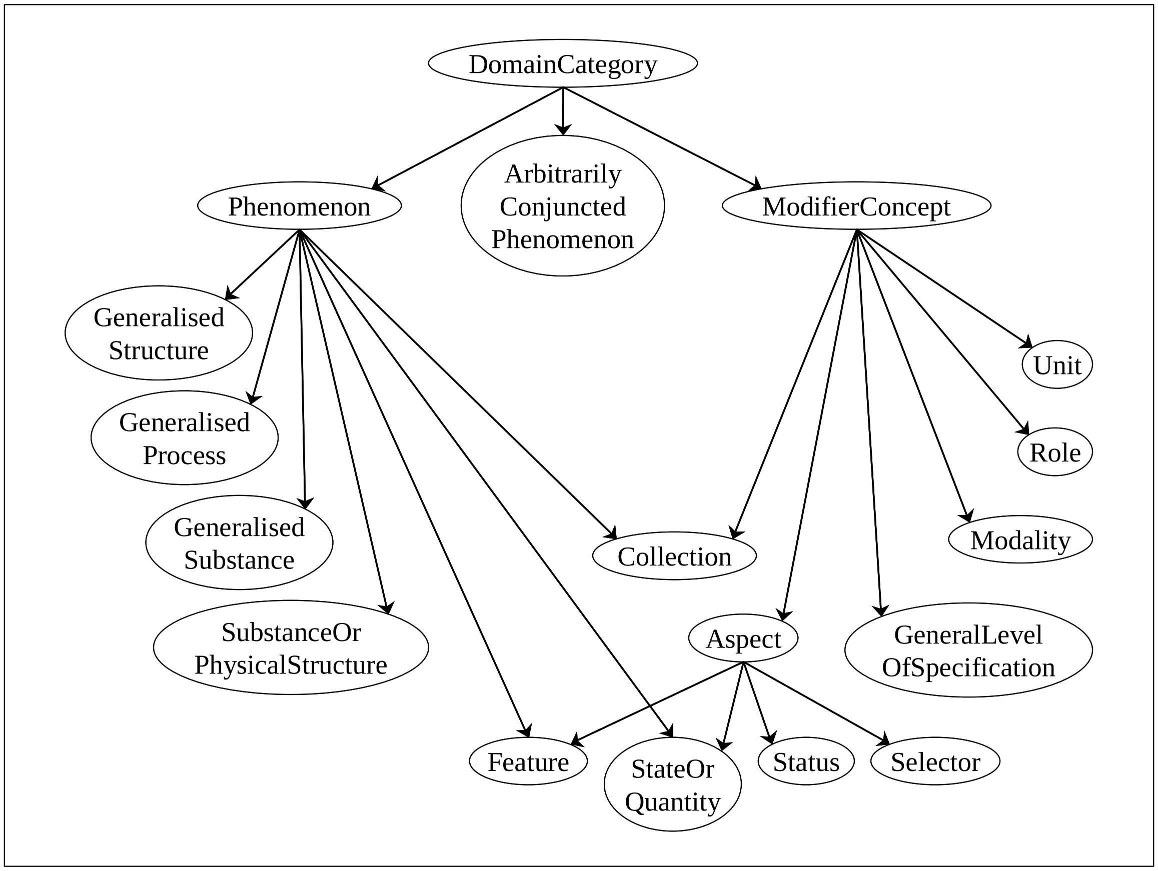

Figure 3 - Top level in OpenGALEN

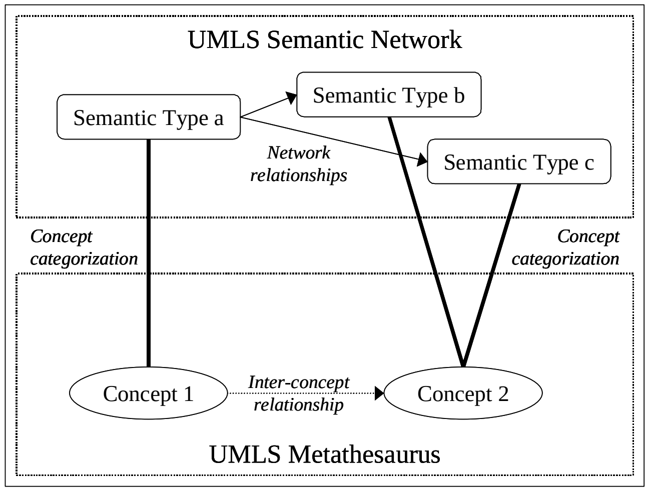

Figure 4 - The two-level structure in the UMLS

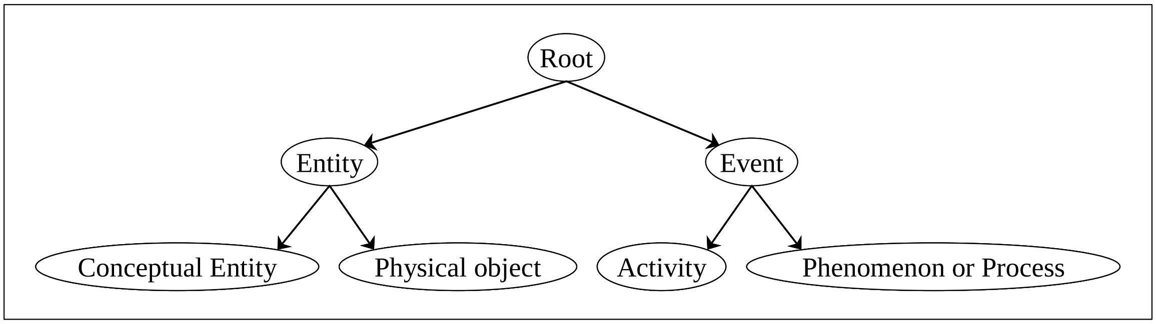

Figure 5 - Top level in the UMLS Semantic Network

Finding, Observation, Clinical history
Disease
Procedure
Body Structure
Anatomical concept
Morphology
Organism
Physical force
Substance
Specimen
Social Context
Attribute
Context dependent categories
Physical object
Event
Environments and geographical location
Observable entity
Qualifier value
Staging and Scales

Figure 6 – Top-level concepts in SNOMED CT

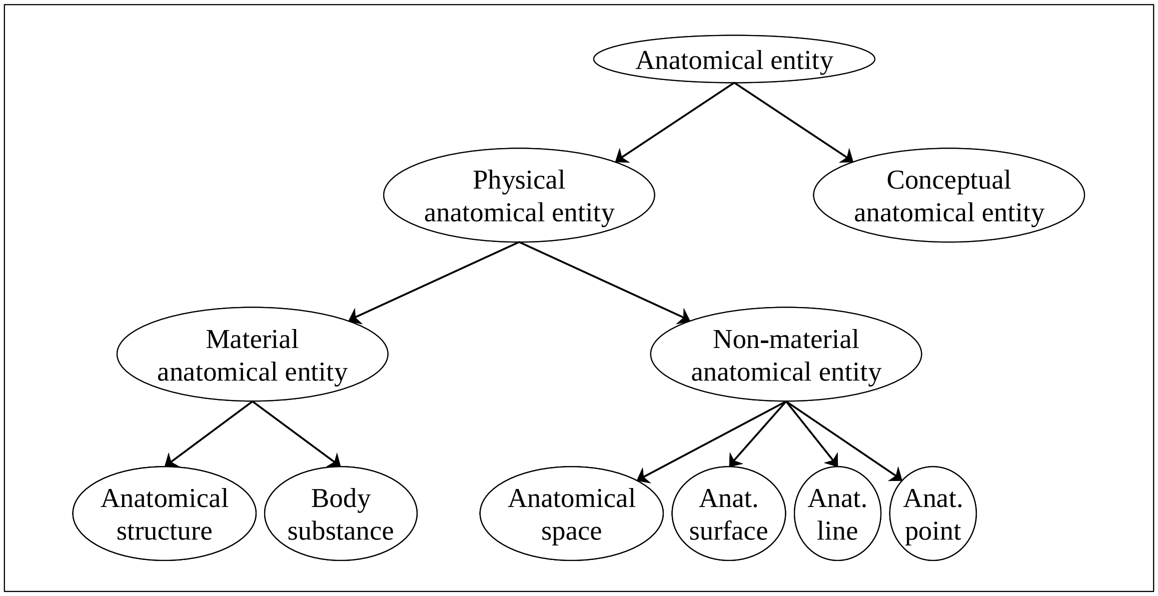

Figure 7 - Top level in Digital Anatomist

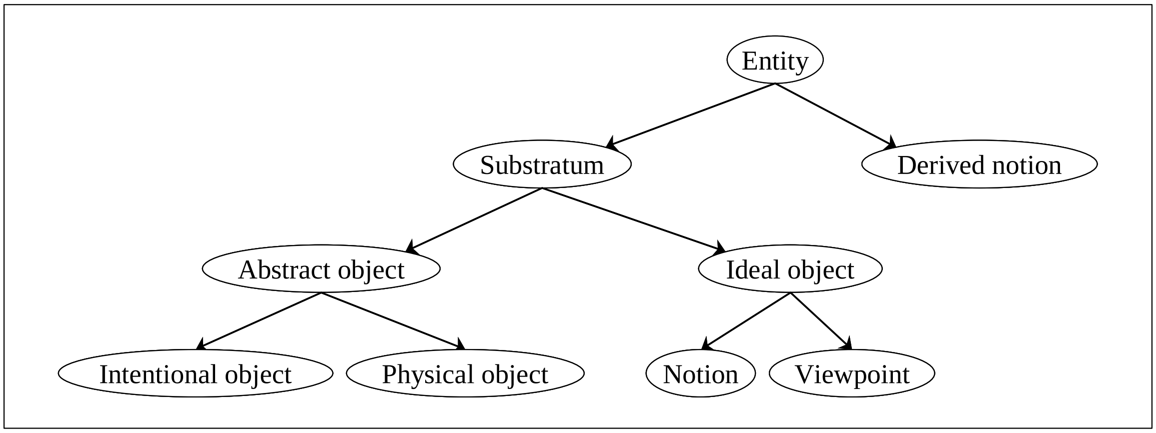

Figure 8 - Top level in MENELAS

a

Blood
- Humor
  - Body fluid
    - Body substance
      - Substance
        - Physical object
          - Entity

a. WordNet
b. OpenGALEN
c. UMLS Semantic Network
d. SNOMED-CT
e. Digital Anatomist
f. MENELAS

d

Blood
- Blood material
  - Body fluid
    - Body substance
      - Human material
        - Biological substance
          - Substance

b

Blood
- SoftTissue
  - Tissue
    - Substance
      - GeneralisedSubstance / SubstanceOrPhysicalStructure
        - Phenomenon
          - DomainCategory

e

Blood
- Body substance
  - Physical anatomical entity
    - Anatomical entity

c

Blood
- Tissue
  - Fully Formed Anatomical Structure
    - Anatomical Structure
      - Physical Object
        - Entity

f

Blood
- Body fluid
  - Tissue
    - Mass object
      - Real object
        - Physical object
          - Abstract object
            - Substratum
              - Entity

Figure 9 - Representation of Blood in several biomedical ontologies

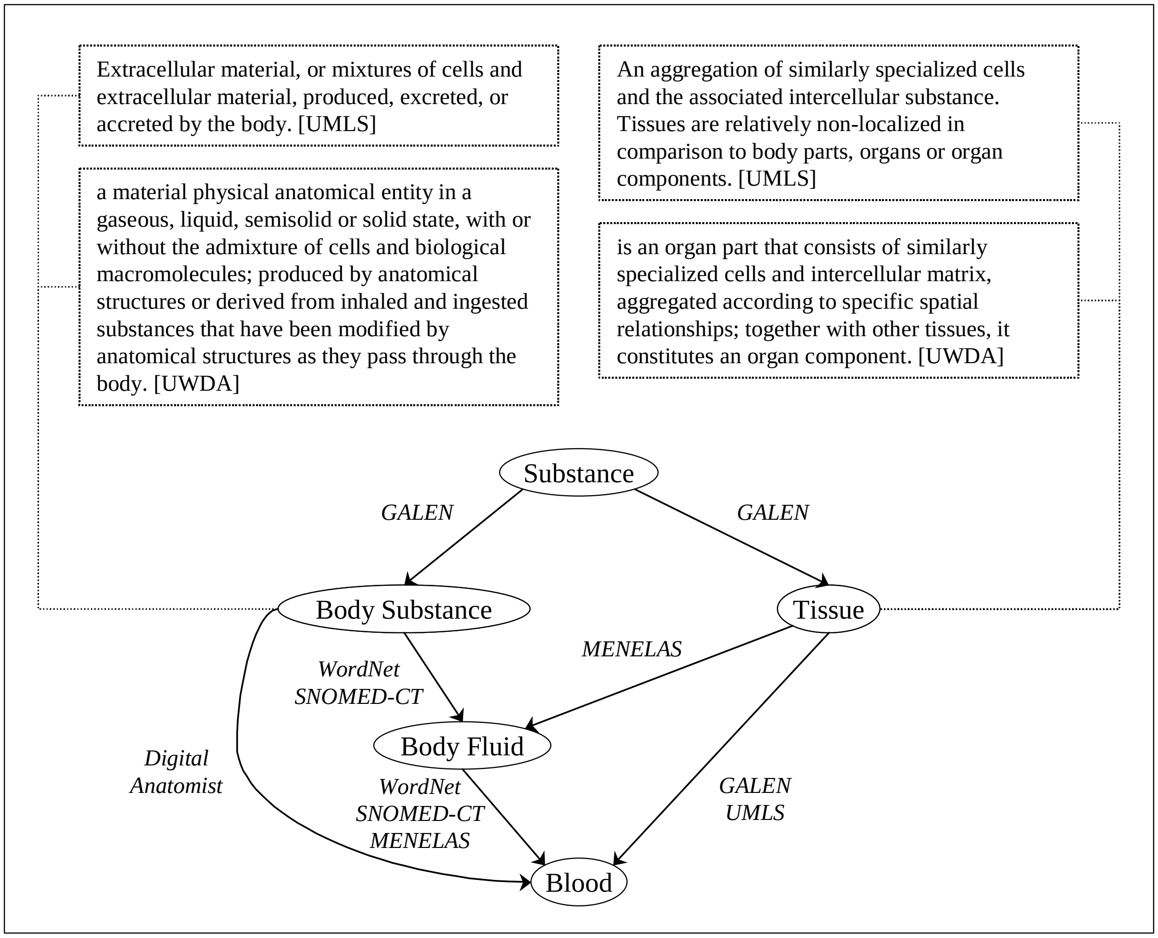

Figure 10 - Composite representation of Blood